DOCUMENT RESUME

ED 100 959 TM 004 047

AUTHOR Benedict, Larry G.
TITLE Traditional Research versus Evaluation.
NOTE 6p.

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Comparative Analysis; Data Collection; *Evaluation; Objectives; Program Evaluation; *Research

ABSTRACT

Research paradigms are not the proper channel for educational evaluation. Evaluation and research differ in many areas, including purpose, methods, goals, groups, and desired outcomes. Research is strictly controlled, has the purpose of gathering information and making generalizations about completed studies or events. Evaluation is a process asking for feedback from groups as they exist, not under controlled conditions. Evaluation seeks specifics that show what is happening in an event while research explains causes. (SM)

ED100959

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

TRADITIONAL RESEARCH versus EVALUATION

Larry G. Benedict
University of Massachusetts

TM 004 047

BEST COPY AVAILABLE

Traditional research paradigms are not adequate for doing educational evaluation. This view is held by Stake, Guba, Stufflebeam and Scriven among others and stems primarily from the fact that both the assumptions and goals of traditional research, perhaps better termed "conclusion-oriented research" (Cronbach and Suppes) are different from those of educational evaluation, which might be termed "decision-oriented research," (Stufflebeam or Cronbach), and thus a paradigm produced on the basis of the assumptions and goals of the former are of necessity and by definition inappropriate in assessing the goals of the latter.

Let's examine briefly some of the assumptions and goals of conclusion-oriented research. First, research has as its primary goal the advancement of knowledge or "Truth." It strives to advance and extend knowledge (Guba). Furthermore, data collected from a research paradigm must be internally valid (Stufflebeam) in order that it be as generalizable as possible (Stake). To achieve all of this a researcher employs the principles of randomization of subjects and treatments, control of extraneous or interacting variables and so on.

However, this is fundamentally different from what educational evaluation strives. Guba says the evaluator is trying to devise and test some practical solution to an operating problem. He is concerned with resolving a number of problems simultaneously if he can. He is concerned also and perhaps most importantly with the need to be able to refine and/or adjust his solutions continuously. Unlike data produced by an experimental design, data which is usually post hoc (Guba; Stufflebeam) evaluation data needs to be continual in order that, as Cronbach points out (and Guba, Stufflebeam and Hastings would all appear to agree) ongoing decisions regarding an educational program may be made while the program is in progress and not after it has been terminated.

BEST COPY AVAILABLE

In fact, according to Stufflebeam, "...the application of experimental design to evaluation problems conflicts with the principle that evaluation should facilitate the continual improvement of a program." (Stufflebeam, p. 49).

Furthermore, evaluation deals with the "worth of something," (Stake) or the "valuing of something," (Scriven) or with "...the use of human judgment," (Glass) and not just the description of something. In the conclusion-oriented paradigms, however, this point is precisely to be avoided at all costs.

Let's also examine more carefully the techniques of research and why they are inadequate for evaluation. Regarding the notion of generalization, there is a basic difference. In fact even the title of Stake's article articulates this difference: "the need for limits." In evaluation, Stake argues, the purpose of inquiry is for "specification" whereas the inquiry in research is for "generalization." He is saying that the purpose of and results of evaluation in fact should not be generalized and cannot be generalized. There is a "need for limits" regarding the generalization of evaluative data. Evaluators are not concerned that findings hold over different schools, over different communities and over replications (Stake). Obviously this is not true of findings in conclusion-oriented research since in order to "extend knowledge" generalizations have to be made, the wider the generalizability, the better.

To achieve control over the threats to validity such as history, maturation, reactive arrangements and so on, the researcher tries to use randomization to assign students to treatment and control groups. He tries to hold all other variables except treatment variables equal during the duration of the experiment. The treatments cannot be modified during the course of the experiment. Again, this is exactly what evaluators do not want and in fact do not and cannot have. Seldom if ever can evaluators exert the kind of control which is demanded by research. (That he doesn't want to is another point.) The evaluator is usually working with a specified problem in a specified setting with specified subjects.

BEST COPY AVAILABLE

He cannot as a rule randomly assign subjects or treatments, run control groups., control for the various threats to validity mentioned in Campbell and Stanley and so on. In addition he does not want to be representative of others, but rather wants to look at the given program for its own value as it is perceived by the decision makers of that program (Guba; Stufflebeam).

Assuming that such tight control can be exerted, as both Guba and Stufflebeam point out, and extraneous variables are held in check, then the findings which result will not even be generalizable to the school or program at hand for in a school or program in the real educational world, these so called extraneous variables operate freely. It is important therefore to know how programs operate under real world conditions and not under the carefully controlled conditions of a laboratory situation (Guba). Stake concurs on this point:

> ....as soon as we exercise a reasonable degree of experimental control, as soon as we provide some variability in the program and hold other aspects constant, the product is altered. Many an educator find the program being researched no longer the program he wanted to know about. (Stake, 1969, p. 2)

There are yet other differences which exclude the utility of experimental designs. Gagne writes that most learning experiments for example have been concerned with the effectiveness of single units of a curriculum, or at the most a very few units. A paradigm such as pre-post test, no control design or a Solomon 4 Block or whatever is fine for examining a single unit, it obviously fails when looking at a larger, ongoing constantly changing program with interacting variables over which there is no control. Stake concludes his argument this way:

> There are two approaches. We have a fundamental choice: to be scientific, to generalize...to find out why; or to be descriptive, to be delimited... to find out what. (Stake, 1969, p. 2)

The former represents conclusion-oriented research, and the latter, evaluation.

BEST COPY AVAILABLE

In summary then I would like to quote from Egon Guba:

...an evaluation paradigm that emphasized control when invited interference is needed; that prevents attention to more than one problem at a time;...that provides only terminal data; and that renders impossible the crucial requirement for continuous adjustment and refinement, simply cannot be judged very useful by the practitioner. Indeed, he must find such a paradigm not only useless but in fact crippling to his purposes. (Guba, 1969, p. 4)

BEST COPY AVAILABLE

References:

Cronbach, Lee J. Evaluation for course improvement. Teacher's College Record, 64:8, 1963, 231-248.

Cronbach, Lee J. and Patrick Suppes, eds. Research for Tomorrow's Schools. Macmillan Co., 1969.

Gagne, Robert. Curriculum research and the promotion of learning. Perspectives of Curriculum Evaluation, AERA Monograph Series, #1, 1969, 19-38.

Glass, Gene. The growth of evaluation methodology. University of Colorado, author, March, 1969.

Guba, Egon. Significant differences. Educational Researcher, XX:3, 1969, 4-5.

Hastings, J. Thomas. Curriculum evaluation: the why of the outcomes. Journal of Educational Measurement, 3:1, 1966, 27-32.

Scriven, Michael. The methodology of evaluation. AERA Monograph, #1, 39-83.

Stake, Robert. Generalizability of program evaluation: the need for limits. Educational Products Report, February, 1969.

Stake, Robert. Toward a technology for the evaluation of educational programs. AERA, #1, 1-12.

Stufflebeam, Daniel. Evaluation as enlightenment for decision making. Improving Educational Assessment, Walcott Beatty, ed., Washington, D.C.: ASCD, 1969, 41-73.